REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Capacity of Generalized Mismatched Gaussian Channels		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL 6. PERFORMING ORG. REPORT NUMBER LAP-2
7. AUTHOR(s) C.R. Baker		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0491 N00014-81-K-0373
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of North Carolina Chapel Hill, North Carolina 27514		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-269 SRO 105
11. CONTROLLING OFFICE NAME AND ADDRESS Statistics & Probability Program Office of Naval Research Arlington, VA 22217		12. REPORT DATE January, 1984 13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release: Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Channel capacity, Gaussian channels, Information theory

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Information capacity is determined for a Gaussian communication channel when the constraint is given in terms of a covariance which is different from that of the channel noise.

LIBRARY
RESEARCH REPORTS DIVISION
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA 93943

CAPACITY OF
GENERALIZED MISMATCHED GAUSSIAN CHANNELS

Charles R. Baker*

Department of Statistics
University of North Carolina University
Chapel Hill, NC 27514

LAP-2

January, 1984

* Research supported by ONR Contracts N00014-75-C-0491 and N00014-81-K-0373

These results were presented at the IEEE International Symposium on Information Theory, St. Jovite, Quebec, Canada, September, 1983.

RUNNING HEAD: Capacity of Gaussian Channels

Author: Charles R. Baker
Department of Statistics
University of North Carolina
Chapel Hill, NC 27514

ABSTRACT

Information capacity is determined for a Gaussian communication channel when the constraint is given in terms of a covariance which is different from that of the channel noise.

AMS (MOS) subject classifications (1970): Primary, 94A15, 60G35

Key words and phrases: Channel capacity, Gaussian channels, information theory.

Introduction

The capacity of the Gaussian channel without feedback, subject to a generalized energy constraint, is determined in [1]. In that work, the constraint is given in terms of the covariance of the channel noise process. However, there are many situations where one may wish to determine capacity subject to a constraint determined by a covariance that is different from that of the channel noise. Examples are jamming or countermeasures situations, or when there is insufficient knowledge of the natural environment.

Channels where the covariance of the noise is the same as that of the constraint will be called matched channels; otherwise, we say that the channel is mismatched (to the constraint). In this paper, the capacity of the mismatched Gaussian channel is determined. Results for a restricted class of mismatched channels are given elsewhere [2]. Various special cases of the mismatched channel have been treated previously [3] - [5].

The results for the mismatched channel differ significantly from those for the matched channel. A discussion of these differences follows the proof of the main result.

Definitions and Structure

The channel is defined as in [1]. H_1 and H_2 are real separable Hilbert spaces with Borel σ fields β_1 and β_2 and inner products $\langle\cdot,\cdot\rangle_1$ and $\langle\cdot,\cdot\rangle_2$. The message process X in H_1 is represented by a probability μ_X on (H_1,β_1). The message is encoded into the transmitted signal A(X) in H_2 by a β_1/β_2-measurable coding function A. To each sample function of the signal process, the channel adds a sample function from the noise process N, represented by a Gaussian measure μ_N on (H_2,β_2). The received signal (channel output) is then a sample function from the process Y = A(X) + N, represented by the measure μ_Y. As usual X and N are assumed independent, so that $\mu_Y(B) = \mu_X \otimes \mu_N\{(X,Y): A(X) + Y \in B\}$ where $\mu_X \otimes \mu_N$ is product measure. The channel probability μ_{XY}, which has marginal measures μ_X and μ_Y, is a measure on $(H_1 \times H_2, \beta_1 \times \beta_2)$ defined by $\mu_{XY}(C) = \mu_X \otimes \mu_N\{(X,Y): (X,A(X) + Y) \in C\}$. The average mutual information is then $I[\mu_{XY}]$, where $I[\mu_{XY}] \equiv \infty$ if it is false that μ_{XY} is absolutely continuous with respect to $\mu_X \otimes \mu_Y$ ($\mu_{XY} \ll \mu_X \otimes \mu_Y$), and otherwise $I[\mu_{XY}] = \int_{H_1 x H_2} \log \left[\frac{d\mu_{XY}}{d\mu_X \otimes \mu_Y}\right](x,y)\, d\mu_{XY}(x,y)$.

The information capacity is then $\sup_Q I[\mu_{XY}]$, where Q is a set of admissible pairs (μ_X, A).

For this paper, a covariance operator in a Hilbert space will be defined to be a symmetric, non-negative, and trace-class bounded linear operator. The constraint on the transmitted signal process A(X) will be given in terms of a covariance operator R_W in H_2; as is well-known, to every such covariance operator there corresponds a zero-mean Gaussian measure on (H_2,β_2).

When $H_2 = L_2[0,T]$ and $\underset{\sim}{R}$ is a covariance operator, $\underset{\sim}{R}$ can be represented as an integral operator with kernel R which is a covariance function. There is then a well-known isomorphism between range $(R^{1/2})$ and the reproducing kernel Hilbert space of R. All measures considered here will be assumed WLOG to have zero mean. The capacity will he determined under the following assumptions:

(A-1) $R_N = R_W^{1/2}(I + S)R_W^{1/2}$, where R_N is the covariance operator of the noise measure μ_N, and S is a bounded linear operator with pure point spectrum that does not have -1 as a limit point of its point spectrum;

(A-2) The admissible set Q is the set of all (μ_X, A) such that $\int_{H_1} ||R_W^{-1/2} A(x)||_2^2 \, d\mu_X(x) \leq P$, where $P > 0$ is fixed.

It will be assumed WLOG [1] that $\overline{\text{range}(R_N)} = H_2$, so that $R_N^{-1/2}$ exists. Assumption (A-1) then implies that $R_W^{-1/2}$ exists; in fact, that range $(R_N^{1/2})$= range $(R_W^{1/2})$. Thus, there exists a unitary operator U in H_2 such that $R_N^{1/2} = R_W^{1/2}(I + S)^{1/2}U^*$, where U^* is the adjoint of U.

The class of all zero-mean Gaussian measures μ_N with covariance operator as in (A-1) includes all those that are mutually absolutely continuous with respect to μ_W, where μ_W is zero-mean Gaussian with covariance R_W[6].

From the results of [1], one can limit attention to cases where $\mu_{A(X)}$ is Gaussian with covariance operator

$$R_{A(X)} = \Sigma_n \tau_n [R_N^{1/2} u_n] \otimes [R_N^{1/2} u_n] \tag{2}$$

where $\tau_n \geq 0$ for $n \geq 1$, $\Sigma_n \tau_n < \infty$, $\{u_n, n \geq 1\}$ is a c.o.n. set and $(u \otimes v)x = \langle v, x \rangle u$.

When $\mu_{A(X)}$ has (2) for covariance and is Gaussian then [1]

$$I[\mu_{XY}] = (1/2) \Sigma_n \log [1 + \tau_n]. \tag{3}$$

Moreover,

$$E_{\mu_X} ||R_W^{-1/2} A(S)||_2^2 = \text{Trace } R_W^{-1/2} R_{A(X)} R_W^{-1/2}$$

$$= \sum_n \tau_n ||(I + S)^{1/2} U^* u_n||_2^2 . \tag{4}$$

Defining $X_n^2 = \tau_n ||(I + S)^{1/2} U^* u_n||_2^2$, the capacity problem is thus reduced to maximizing

$$(1/2) \sum_n \log [1 + X_n^2 (1 + \gamma_n)^{-1}] \tag{5}$$

over all sequences (X_n^2) and c.o.n. sets $\{v_n, n \geq 1\}$ such that $\Sigma_n X_n^2 \leq P$, where $\gamma_n \equiv \langle Sv_n, v_n \rangle_2$, $n \geq 1$.

The suprem um of (5) subject to the stated constraint is the capacity and will be denoted as $C_W(P)$; the capacity for the matched channel ($R_W = R_N$) will be denoted by $C_N(P)$.

In the case where H_2 is infinite-dimensional, θ will denote lim inf of the eigenvalues of the operator S satisfying $R_N = R_W^{1/2}(I+S)R_W^{1/2}$. By assumption (A-1), $\theta > -1$. $\{\lambda_n, n\geq 1\}$ will denote the eigenvalues of S that are strictly less than θ; of course, this set can be empty. Similarly, $\{e_n, n\geq 1\}$ will always denote an o.n. set of H_2 elements that are eigenvectors of S corresponding to the eigenvalues $\{\lambda_n, n\geq 1\}$: $Se_n = \lambda_n e_n$, $n\geq 1$.

Preliminary Results

Lemma 1: Let (γ_n), $n \leq M$, be any non-decreasing sequence of strictly positive real numbers. Let (X_n) be any sequence of M real numbers. Fix $P > 0$, and define

$$g(M,P,\underset{\sim}{\gamma}) = \sup_{\{\underset{\sim}{X}:\ \sum_1^M X_n^2 \leq P\}} \prod_1^M (\gamma_n + X_n^2)/\gamma_n.$$

Then $g(M,P,\underset{\sim}{\gamma}) = \prod_{n=1}^{K} (\sum_1^K \gamma_i + P)/(K\gamma_n)$

where $K \leq M$ is the largest integer such that $\sum_1^K \gamma_i + P \geq K\gamma_K$. $g(M,P,\underset{\sim}{\gamma})$ is uniquely attained by (X_n^2) such that

$$X_n^2 = (\sum_1^K \gamma_i + P)/K - \gamma_n \qquad n \leq K$$

$$= 0 \qquad n > K.$$

Proof: Define f_M: $\mathbb{R}^M \to \mathbb{R}$ by $f_M(\underline{y}) = \sum_{n=1}^{M} \log [1 + y_n \gamma_n^{-1}]$. We seek to maximize f_M subject to the constraints

$$g(\underline{y}) = \sum_1^M y_n - P \leq 0$$

$$h_i(\underline{y}) = -y_i \leq 0, \quad i = 1,.., M.$$

This is a constrained optimization problem with objective function f_M which is concave over the convex set $\{\underline{Z} \text{ in } \mathbb{R}^M;\ Z_i \geq 0,\ i = 1,..,M\}$. Moreover, each constraint function is linear. Thus, any solution to this problem will define a unique global maximum for f_M [7]. In order that $\underline{y}^*$ be a solution, it is necessary and sufficient that the following set of equations be satisfied [7]:

$$\frac{1}{y_i^* + \gamma_i} + \beta - \alpha_i = 0 \qquad i = 1,.., M \tag{6}$$

$$\sum_1^M y_n^* - P \le 0, \quad \beta[\sum_1^M y_n^* - P] = 0 \tag{7}$$

$$-y_i^* \le 0, \quad \alpha_i y_i^* = 0, \quad i = 1,..,M \tag{8}$$

for some set of non-positive real numbers $\{\beta, \alpha_1,..,\alpha_M\}$.

We first attempt to obtain a solution by setting $\alpha_1 = \alpha_2 = \ldots = \alpha_M = 0$. This requires $\beta(\gamma_i + y_i^*) = -1$ for $i = 1,..,M$; thus,

$$\sum_1^M y_i^* + \sum_1^M \gamma_i = -M\beta^{-1}, \text{ and so } y_n^* = (\sum_1^M y_i^* + \sum_1^M \gamma_i)/M - \gamma_n$$

for $n = 1, 2,..,M$. This definition of $\underset{\sim}{y}^*$ and the constraints (8) require that

$$\sum_1^M y_i^* + \sum_1^M \gamma_i \ge M\gamma_n$$

for $n \le M$; this inequality is satisfied for all $n \le M$ if and only if it is satisfied for $n = M$. Also, $\beta^{-1} = -(y_i^* + \gamma_i)$ for $i \le M$ implies $\beta < 0$, so that $\sum_1^M y_i^* = P$ by constraints (7). Thus, if $P + \sum_1^M \gamma_i \ge M\gamma_M$, an optimum solution is given by

$$y_i^* = (P + \sum_1^M \gamma_n - M\gamma_i)/M, \qquad i \le M.$$

If there exists $K<M$ such that $K\gamma_K \le P + \sum_1^K \gamma_i < (K+1)\gamma_{K+1}$,

then constraints (6)-(8) are satisfied by choosing $\beta = -K[P + \sum_1^K \gamma_i]^{-1}$,

$$\alpha_1 = \alpha_2 = .. = \alpha_K = 0,$$

$$\sum_1^K y_i^* = P$$

$$y_i^* = 0, \ i > K$$

$$y_i^* = K^{-1}[P + \sum_1^K \gamma_n - K\gamma_i], \quad i \le K$$

$$\alpha_i = -K[P + \sum_1^K \gamma_n]^{-1} + \gamma_i^{-1} \quad i > K$$

Thus,

$$\sup_{\{\underset{\sim}{X}:\ \sum_1^M X_n^2 \le P\}} \prod_{n=1}^M (\gamma_n + X_n^2)/\gamma_n = \prod_{n=1}^K (\sum_1^K \gamma_i + P)/(K\gamma_n)$$

where $K \leq M$ is the largest integer such that $\sum_1^K \gamma_i + P \geq K\gamma_K$. The supremum is attained by $\underline{y}^*$ as defined above, or for

$$x_n^2 = [P + \sum_1^K \gamma_i]/K - \gamma_n \qquad n \leq K$$

$$= 0 \qquad n > K.$$

□

Lemma 2: Let (λ_i), $1 \leq i \leq K$, be a non-decreasing sequence of strictly positive real numbers. Suppose that (γ_n) is a non-decreasing sequence such that $\sum_1^J \gamma_i \geq \sum_1^J \lambda_i$ for all $J \leq K$, and let $P > 0$ be such that $\sum_1^K \gamma_i + P \geq K\gamma_K$. Define

$f_K(\underset{\sim}{\gamma}) = \prod_{n=1}^K (P + \sum_1^K \gamma_i)/(K\gamma_n)$. Then $f_K(\underset{\sim}{\gamma}) \leq f_K(\underset{\sim}{\lambda})$ with equality if and only if $\gamma_i = \lambda_i$ for all $i \leq K$.

Proof: For any fixed n, $\partial f_K(\gamma)/\partial\gamma_n$ is negative, using $\sum_1^K \gamma_i + P \geq K\gamma_K$. Thus $f_K(\underset{\sim}{\gamma})$ increases for γ_n decreasing. One can now assume that $\sum_1^K \gamma_n = \sum_1^K \lambda_n$. To see this, suppose $\sum_1^K \gamma_n > \sum_1^K \lambda_n$. First assume that there exists $p \leq K$ such that $\gamma_p > \gamma_{p-1}$ and $\gamma_p > \lambda_p$. Define a sequence (γ_n') by $\gamma_n' = \gamma_n$ if $n \neq p$, $\gamma_p' = \gamma_p - \varepsilon$,

$$\varepsilon = \min\,(\gamma_p - \gamma_{p-1}, \sum_1^K (\gamma_i - \lambda_i), \gamma_p - \lambda_p).$$

Continuing to form new sequences in this manner, one will eventually obtain a non-decreasing sequence (γ_n') with $\sum_1^J \gamma_n' \geq \sum_1^J \lambda_n$ for all $J \leq K$, and either $\sum_1^K \gamma_n' = \sum_1^K \lambda_n$ or else $\gamma_1' = \gamma_2' = \ldots = \gamma_p'$, where p is the largest integer i such that $\gamma_i > \lambda_i$. If the latter case holds, define a new sequence (γ_n''), with $\gamma_n'' = \gamma_n'$ for $2 \leq n \leq K$, while $\gamma_1'' = \gamma_1' - \varepsilon$, $\varepsilon = \min\,(\gamma_1' - \lambda_1, \sum_1^K(\gamma_i' - \lambda_i))$. (γ_n'') is non-decreasing and $\sum_1^J \gamma_n'' \geq \sum_1^J \lambda_n$. If $\varepsilon = \gamma_1' - \lambda_1$, the procedure is repeated for (γ_n'') and γ_2''; if $\varepsilon = \sum_1^K(\gamma_i' - \lambda_i)$, the procedure is repeated for (γ_n'') and γ_1''. Continuing in this manner will eventually produce a sequence (γ_n'') such that $\sum_1^K \gamma_n'' = \sum_1^K \lambda_n$.

Assume then that $\sum_1^K \gamma_n = \sum_1^K \lambda_n$. If $\underset{\sim}{\gamma}$ and $\underset{\sim}{\lambda}$ are not identical, let $p \leq K$ be the largest integer such that $\gamma_p \neq \lambda_p$; since $\sum_1^{p-1} \gamma_n \geq \sum_1^{p-1} \lambda_n$ and $\sum_1^{p} \gamma_n = \sum_1^{p} \lambda_n$, $\gamma_p < \lambda_p$. Let $t < p$ be the largest integer such that $\gamma_t > \lambda_t$; such t must exist. Define a new sequence (γ_n') by $\gamma_n' = \gamma_n$ if $n \neq t$, $n \neq p$, while $\gamma_p' = \gamma_p + \varepsilon$, $\gamma_t' = \gamma_t - \varepsilon$, $\varepsilon = \inf(\lambda_p - \gamma_p, \ \gamma_t - \lambda_t)$. $f_K(\underset{\sim}{\gamma}) < f_K(\underset{\sim}{\gamma}')$, since $(\gamma_t - \varepsilon)(\gamma_p + \varepsilon) = \gamma_t\gamma_p - \varepsilon(\gamma_p - \gamma_t) - \varepsilon^2$, and $\gamma_p \geq \gamma_t$. This procedure is successively repeated; it will terminate when and only when one obtains a sequence (γ_n') such that $\gamma_n' = \lambda_n$ for all $n \leq K$.

$\square$

Main Results

Theorem 1:

(a) Suppose that H_2 has dimension $M < \infty$. The capacity is then

$$C_W(P) = (1/2) \sum_{n=1}^{K} \log \left[\frac{\sum_1^K \beta_i + P + K}{K(1 + \beta_n)} \right]$$

where $\beta_1 \leq \beta_2 \leq .. \leq \beta_M$ are the eigenvalues of S, and K is the largest integer $\leq M$ such that $\sum_1^K \beta_i + P \geq K\beta_K$. The capacity is attained by a Gaussian $\mu_{A(X)}$ with covariance operator (2), where $u_n = Ue_n$ and $\tau_n = \left[\sum_1^K \beta_i + P - K\beta_n \right] (1 + \beta_n)^{-1} K^{-1}$ for $n \leq K$, $\tau_n = 0$ for $n > K$, and $\{e_n, n \geq 1\}$ are o.n. eigenvectors of S corresponding to the eigenvalues (β_n). No other Gaussian $\mu_{A(X)}$ can attain capacity. The same result is obtained if H_2 has dimension $L < \infty$ and $\mu_{A(X)}$ is constrained to have support of dimension $M < L$.

(b) Suppose that H_2 is infinite-dimensional and that support $(\mu_{A(X)})$ is restricted to have dimension $\leq M < \infty$.

(i) If $\{\lambda_n, n\geq 1$ is empty, then $C_W(P) = (M/2) \log [1 + PM^{-1}(1+\theta)^{-1}]$. Capacity can be attained if and only if S has θ as an eigenvalue of multiplicity $\geq M$. In this case $C_W(P)$ is attained by a Gaussian $\mu_{A(X)}$ with covariance (2), where $u_i = Ue_i$ and $\tau_i = PM^{-1}(1+\theta)^{-1}$ for $i\leq M$, with $\{e_1,..,e_M\}$ any o.n. set in the null space of S.

(ii) If $K\lambda_K \leq \sum_1^K \lambda_i + P < K\lambda_{K+1}$ for some $K<M$, then the capacity is as in (a), with $\beta_i = \lambda_i$, $i = 1,..,K$, and can be similarly attained.

(iii) Let $K = \min(L,M)$, where $L \geq 1$ is the number of eigenvalues of S whose value is strictly less than θ, and suppose that $P + \sum_1^K \lambda_i \geq K\lambda_K$. The capacity is then

$$C_W(P) = \tfrac{1}{2}\sum_{n=1}^{K} \log \left[\frac{1+\theta}{1+\lambda_n}\right] + \left(\frac{M}{2}\right) \log \left[1 + \frac{P + \sum_{i=1}^{K}(\lambda_i - \theta)}{M(1+\theta)}\right].$$

The capacity can be attained if and only if θ is an eigenvalue of S with multiplicity $\geq$ M-K. The capacity is then achieved by a Gaussian μ_{AX} with covariance (2), where $u_n = Ue_n$ and $\tau_n = (\sum_1^K \lambda_i + P - M\lambda_n - K\theta)(1 + \lambda_n)^{-1}M^{-1}$ for $n \leq K$, with $Se_n = \lambda_n e_n$ and $\{e_1,..,e_K\}$ an o.n. set; and with $u_n = Uv_n$ and $\tau_n = (P+\sum_1^K \lambda_i - K\theta)M^{-1}(1+\theta)^{-1}$ for $K + 1 \leq n \leq M$, where $Sv_n = \theta$ and $v_{K+1},..,v_M$ is an o.n. set. The sets $\{u_1,..,u_K\}$ and $\{\tau_1,..,\tau_K\}$ are uniquely defined for any maximizing Gaussian $\mu_{A(X)}$.

<u>Proof</u>: (a). From (5),

$C_W(P) = \sup \left[\tfrac{1}{2}\sum_1^M \log [1 + X_n^2\lambda_n^{-1}]\right]$, where $\gamma_n = 1 + \langle Sv_n, v_n\rangle_2$, $\{v_n, n\leq M\}$ is a c.o.n. set, and the supremum is over all such c.o.n. sets and all (X_n^2) such that $\sum_1^M X_n^2 \leq P$. Since $\beta_1 \leq \beta_2 \leq .. \leq \beta_M$ are the non-decreasing eigenvalues of S,

$\sum_1^J [1 + \langle Sv_n, v_n\rangle_2] \geq \sum_1^J [1 + \beta_n]$ for all $J \leq M$ and any fixed c.o.n. set $\{v_n, n \leq M\}$. The expression of $C_W(P)$ in (a) and the unique covariance of the maximizing Gaussian $\mu_{A(X)}$ both now follow from Lemma 1 and Lemma 2. The same result holds if $\dim (H_2) = L < \infty$ and $\dim [\mathrm{supp}(\mu_{A(X)})] \leq M < L$, since in this case S again has M smallest eigenvalues.

(b) - (i). If $S > \theta I$, then $S-\theta I$ does not have M smallest eigenvalues. However, given any $\varepsilon > 0$, one can find eigenvalues $\gamma_1^\varepsilon, \ldots, \gamma_M^\varepsilon$ such that $\theta < \gamma_i^\varepsilon < \theta + \varepsilon$ for $i \leq M$. Using this in (3) one obtains

$$I[\mu_{XY}] = (1/2) \sum_1^M \log [1 + \tau_n]$$

$$= (1/2) \sum_1^M \log [1 + X_n^2 (1 + \gamma_n^\varepsilon)^{-1}] \geq (1/2) \sum_1^M \log [1 + X_n^2 (1 + \theta + \varepsilon)^{-1}].$$

The expression on the right of the inequality is maximized, over all (X_n^2) such that $\sum_1^M X_n^2 \leq P$, by defining $X_n^2 = P/M$, $n \leq M$. Thus, $C_W(P) \geq (1/2) \sum_1^M \log [1 + (1 + \theta + \varepsilon)^{-1} P/M]$ for all $\varepsilon > 0$, and so $C_W(P) \geq (1/2) \sum_1^M \log [1 + PM^{-1}(1 + \theta)^{-1}]$. For the reverse inequality, one notes that under the constraint $E_{\mu_X} ||R_N^{-1/2} A(X)||_2^2 \leq P$, it is shown in [1] that $C_N(P) = (M/2) \log (1 + P/M)$. For $S \geq \theta I$, $||R_N^{-1/2} A(X)||_2^2 \leq (1+\theta)^{-1} ||R_W^{-1/2} A(X)||_2^2$. Thus, $E_{\mu_X} ||R_W^{-1/2} A(X)||_2^2 \leq P$ implies $E_{\mu_X} ||R_N^{-1/2} A(X)||_2^2 \leq (1+\theta)^{-1} P$, giving $C_W(P) \leq \left(\frac{M}{2}\right) \log [1 + \frac{P}{M(1+\theta)}]$, so that $C_W(P) = (M/2) \log [1 + PM^{-1}(1+\theta)^{-1}]$.

If $S \geq \theta I$, with θ an eigenvalue of multiplicity K, the above argument is

modified in an obvious way ($\gamma_1^{\varepsilon} = \theta$ for $i = 1,\ldots,\min(K,M)$) to again obtain $C_W(P) = (M/2)\log[1 + PM^{-1}(1+\theta)^{-1}]$.

To prove (b-ii), the proof of (a) is repeated after substituting λ_i for β_i, $i \leq M$.

Now suppose that S has $K < M$ strictly negative eigenvalues $\lambda_1 \leq \cdots \leq \lambda_K$, and that $\sum_1^K \lambda_i + P \geq K\lambda_K$. $C_W(P) = \sup_{(P_1,\underline{v})} C_W(P_1,\underline{v})$ where

$$C_W(P_1,\underline{v}) = \sup\left[\frac{1}{2}\sum_{n=1}^{M} \log\left[1 + X_n^2\,(1 + \langle Sv_n, v_n\rangle_2)^{-1}\right]\right],$$

$\underline{v} = \{v_n, n \leq M\}$ is any o.n. set, $0 \leq P_1 \leq P$, and the supremum is over all (X_n^2) such that $\sum_1^K X_n^2 \leq P_1$, $\sum_1^M X_n^2 \leq P$. Repeating the analysis of (a) and (b-i), one finds that

$$C_W(P_1,\underline{v}) = (1/2)\sum_{n=1}^{J} \log\left[\frac{\sum_1^J \lambda_i + P_1 + J}{J(1+\lambda_n)}\right]$$

$$+ (1/2)(M-K)\log\left[1 + \frac{P - P_1}{(M-K)(1+\theta)}\right]$$

where $J \leq K$ is the largest integer such that $\sum_1^J \lambda_i + P_1 \geq J\lambda_J$. Since this result holds for any o.n. set $\{v_n, n \leq M\}$, it remains only to determine the value of P_1 that maximizes $C_W(P_1,\underline{v})$ (a differentiable function of P_1 in $[0,P]$). Differentiating, one sees that $C_W(P_1,\underline{v})$ is increasing with P_1 so long as $P_1 < [JP + (M-K)(J\theta - \sum_1^J \lambda_i)](M-K+J)^{-1}$. Since $P_1 < J\lambda_{J+1} - \sum_1^J \lambda_i$, the preceding inequality is satisfied so long as $(M-K+J)\lambda_{J+1} - \sum_1^J \lambda_i < P + (M-K)\theta$ and this is satisfied because $P + \sum_1^J \lambda_i \geq J\lambda_{J+1}$, $\lambda_{J+1} < \theta$. It follows that $C_W(P_1,\underline{v})$ is an

increasing function of P_1 for $P_1 < -\sum_1^K \lambda_i + K\lambda_K$. Assuming that $P_1 \geq -\sum_1^K \lambda_i + K\lambda_K$, the maximum of $C_W(P_1, \underline{v})$ is attained uniquely by $P_1 = M^{-1}[KP - (M-K)\sum_1^K \lambda_i + (M-K)K\theta]$. Using this value of P_1 in the expression for $C_W(P_1, \underline{v})$, one obtains $C_W(P)$ as in (b-iii). The statement on attaining capacity follows from the results of (b-i) and (b-ii). □

Theorem 2: Suppose that H_2 is infinite-dimensional, and that $\dim[\mathrm{supp}(\mu_{AX})]$ is not constrained.

(a) If $\{\lambda_n, n\geq 1\}$ is not empty, and $\sum_n (\theta-\lambda_n) \leq P$, then

$$C_W(P) = \frac{1}{2}\sum_n \log\left[\frac{1+\theta}{1+\lambda_n}\right] + \frac{1}{2}\,\frac{P + \sum_m(\lambda_m - \theta)}{1+\theta}.$$

(b) If $\{\lambda_n, n\geq 1\}$ is not empty, and $P < \sum_n (\theta - \lambda_n)$, then there exists a largest integer K such that $\sum_1^K \lambda_i + P \geq K\lambda_K$, and $C_W(P) = \frac{1}{2}\sum_{n=1}^K \log\left[\frac{\sum_1^K \lambda_i + P + K}{K(1+\lambda_n)}\right]$.

(c) If $\{\lambda_n, n\geq 1\}$ is empty, then $C_W(P) = \frac{P}{2(1+\theta)}$.

(d) In (a), the capacity can be attained if and only if $\sum_n (\theta-\lambda_n) = P$. It is then attained by a Gaussian μ_{AX} with covariance operator as in (2), where $u_n = Ue_n$ and $\tau_n = (\theta-\lambda_n)(1+\lambda_n)^{-1}$ for all $n\geq 1$. In (b), the capacity can be attained by a Gaussian μ_{AX} with covariance operator (2), where

$$u_n = Ue_n \text{ and } \tau_n = \frac{\sum_1^K \lambda_i + P + K}{K(1+\lambda_n)} - 1 \text{ for } n\leq K;\ \tau_n = 0 \text{ for } n>K.$$

In (c), the capacity cannot be attained.

Remarks. Theorem 2 has been proved for the case $\theta=0$ in [2]. Here we shall omit the detailed proof of Theorem 2 above; given lemmas 3-5 below, the proof of the

Theorem in [2] can be applied in exactly the same way to prove the present Theorem 2.

Several errata in [2] should be mentioned here. The expression for τ_n $(n \le K)$ in the first line of p. 9 of [2] should read:

$\tau_n = (1+\lambda_n)^{-1}\left[\sum_1^K \lambda_i + P - K\lambda_n\right]/K$. The expression for τ_n on the next-to-last line of p. 8 should read: $\tau_n = -\lambda_n(1+\lambda_n)^{-1}$ for $n \ge 1$. On p. 11, sixth line from bottom, P_i should be P_1; in the fifth line from bottom, $C(P_1, K)$ should read

$C(P_1,K) = \frac{1}{2}\sum_{n=1}^{K} \log\left[\frac{\sum_1^K \lambda_i + P_1 + K}{K(1+\lambda_n)}\right] + \frac{1}{2}(P - P_1)$. These are all typing errata and do not affect the development in [2].

Lemma 3. Suppose that $\{\lambda_n, n \ge 1\}$ is an infinite set, and $P>0$. Then $P + \sum_1^K \lambda_n \ge K\lambda_K$ for all $K \ge 1$ if and only if $P \ge \sum_{n\ge 1}(\theta - \lambda_n)$.

Proof. $P + \sum_1^K (\lambda_n - \theta) \ge K(\lambda_K - \theta) \iff P + \sum_1^K \lambda_n \ge K\lambda_K$; the result follows from Lemma 3 of [2].

Lemma 4. Suppose that $S - \theta I$ is strictly negative.

(a) If $P \ge \sum_n (\theta - \lambda_n)$, then $C_W(P) = \frac{1}{2}\sum_n \log\left[\frac{1+\theta}{1+\lambda_n}\right] + \frac{1}{2}\left[\frac{P + \sum(\lambda_n - \theta)}{1+\theta}\right]$.

The capacity can be attained can be attained if and only if $P = \sum_n (\theta - \lambda_n)$. It is then attained by a Gaussian μ_{AX} with covariance operator (2), where $u_n = Ue_n$ and $\tau_n = (\theta - \lambda_n)(1+\lambda_n)^{-1}$ for all $n \ge 1$.

(b) If $P < \sum_n (\theta - \lambda_n)$, then $C_W(P) = \frac{1}{2}\sum_{n=1}^{K} \log\left[\frac{\sum_1^K \lambda_i + P + K}{K(1+\lambda_n)}\right]$ where $K<\infty$ is the largest integer such that $P + \sum_1^K \lambda_n \ge K\lambda_K$. The capacity can be attained by

a Gaussian μ_{AX} with covariance operator (2), where $u_n = Ue_n$ and

$$\tau_n = \frac{\sum_1^K \lambda_i + P + K}{K(1+\lambda_n)} - 1 \quad \text{for } n \leq K$$

$$\tau_n = 0 \qquad\qquad \text{for } n > K.$$

Proof. (a). The fact that

$$2C_W(P) \geq \sum_{n\geq 1} \log \left[\frac{1+\theta}{1+\lambda_n}\right] + \frac{P + \sum_m(\lambda_m - \theta)}{1+\theta} \tag{*}$$

follows from b-(iii) of Theorem 1, letting $M \to \infty$ in that result. To prove the reverse inequality, one repeats the proof of part (a) of Lemma 4 in [2], substituting the RHS of (*) above for the RHS of (*) in [2 , p.11]. For this, one uses the fact that

$$\sum_{n=1}^M \log \left[\frac{\sum_1^M \lambda_i + P + M}{M(1+\lambda_n)}\right] = \sum_{n=1}^M \log \left[\frac{1+\theta}{1+\lambda_n}\right] + M \log \left[1 + \frac{\sum_{i=1}^M (\lambda_i - \theta) + P}{M(1+\theta)}\right].$$

The result on attaining capacity is proved in exactly the same way as the corresponding result in [2], again after substituting the RHS of (*) above for the RHS of (*) in [2, p.11].

$$\text{(b). } C_W(P) \geq \frac{1}{2} \sum_{n-1}^K \log \left[\frac{\sum_{i=1}^K \lambda_i + P + K}{K(1+\lambda_n)}\right] \tag{**}$$

follows from (b-ii) of Theorem 1. Suppose that $C_W(P) >$ RHS(**). Then by (5) there exists a c.o.n. set $\{v_n, n\geq 1\}$ and a sequence (X_n^2) with an infinite number of non-zero terms (using (b-ii) of Theorem 1) such that

$$\text{RHS}(**) < \frac{1}{2} \sum_n \log [1 + X_n^2(1 + \langle Sv_n, v_n\rangle_2)^{-1}] \tag{***}$$

with $\sum_n X_n^2 \leq P$. As $\sum_{n=1}^\infty X_n^2 < \infty$ and $(1 + \langle Sv_n, v_n\rangle)$ is a bounded sequence,

(***) implies that for some $M<\infty$, RHS(**) $< \frac{1}{2} \sum_{n=1}^M \log [1 + X_n^2(1 + \langle Sv_n, v_n\rangle_2)^{-1}]$.

This contradicts b-(ii) of Theorem 1. □

Lemma 5. If $S \geq \theta I$, then $C_W(P) = \frac{P}{2(1+\theta)}$.

Proof. $C_W(P) \geq P(1+\theta)^{-1}2^{-1}$ follows from part b-(i) of Theorem 1, by letting $M \to \infty$. To prove the reverse inequality, one notes that for the constraint $E_{\mu_X}||R_N^{-1/2} A(X)||_2^2 \leq (1+\theta)^{-1}P$, the capacity $(C_N[(1+\theta)^{-1}P])$ is $P(1+\theta)^{-1}2^{-1}$ [1, Theorem 2]. Since $E_{\mu_X}||R_W^{-\frac{1}{2}} A(X)||_2^2 \leq P$ implies $E_{\mu_X}||R_N^{-1/2} A(X)||_2^2 \leq (1+\theta)^{-1}P$, optimization w.r.t. the former constraint is over a smaller set than w.r.t. the latter constraint; thus $C_W(P) \leq C_N[(1+\theta)^{-1}P]$. □

Comparisons of $C_W(P)$ and $C_N(P)$

For the finite-dimensional channel, the capacity $C_W(P)$ given in Theorem 1(a) is strictly greater than $C_N(P)$ $(= \frac{1}{2} \log [1+P/M])$ if $\sum_1^M \beta_i \leq 0$, or if $P + \sum_1^K \beta_i \leq 0$. $C_W(P) < C_N(P)$ if $0 \leq \beta_1 < \beta_M$. The verification is omitted.

For the infinite-dimensional channel, a general statement can be made if $\{\lambda_n, n\geq 1\}$ is empty. Then, $C_W(P) > C_N(P)$ if $\theta<0$, $C_W(P) < C_N(P)$ if $\theta>0$, $C_W(P) = C_N(P)$ if $\theta=0$; see Theorem 1 (b-i) and Theorem 2 (c). Note that $C_N(P) = P/2$ for the unconstrained channel [1, Theorem 2].

If $\{\lambda_n, n\geq 1\}$ is not empty, then for the unconstrained channel the value of $C_W(P)$ given in Theorem 2(a) is greater than $\frac{P}{2(1+\theta)}$, using $\log x^{-1} \geq 1-x$. This inequality can also be shown for the value given in Theorem 2(b), proceeding as in the proof of part (b) of the Theorem in [2]. Thus, for the unconstrained channel, $C_W(P) > C_N(P)$ if $\theta\leq 0$ and $\{\lambda_n, n\geq 1\}$ is not empty. A similar result can be obtained for the constrained channel.

Discussion.

The mismatched channel differs from the matched channel in several ways. First, the value of the capacity can be very different, as already seen. Secondly, the problem of attaining capacity is much more significant. Even in the finite-dimensional channel the vectors $u_1,..,u_M$ must be a specific set of vectors, not just any o.n. set. If H_2 is infinite-dimensional with $\dim[\mathrm{supp}(\mu_{A(X)})] \leq M$, the situation is even worse in (b-iii) of Theorem 1. That is, capacity can then be attained only if S has zero as an eigenvalue of multiplicity $\geq M$ when $S \leq \theta I$, or of multiplicity $\geq M-K$ when S has $K < M$ strictly negative eigenvalues $\lambda_1 \leq .. \leq \lambda_K$ and $P + \sum_1^K \lambda_i \geq K\lambda_K$. Otherwise, in order to approach capacity, one will need to put part of the available "energy" P in elements (Ue_n) where (e_n) are eigenvectors of S corresponding to successively smaller eigenvalues. In practical applications, this usually corresponds to eigenfunctions at higher and higher frequencies.

For the infinite-dimensional channel without a constraint on $\dim[\mathrm{supp}(\mu_{AX})]$, again there can be significant differences between $C_W(P)$ and $C_N(P)$, depending on $\{\theta;\ \lambda_n,\ n\geq 1\}$. However, in this case one sees a rather different situation in the problem of attaining capacity. $C_N(P)$ can never be attained; $C_W(P)$ can be attained if and only if $\{\lambda_n, n\geq 1\}$ is not empty and $P \leq \sum_n (\theta-\lambda_n)$.

It may be noted that the results given in (a) and (b-ii) of Theorem 1 are similar to those obtained in [4, p. 170], although the developments are quite different. However, those previous results are given in terms of a constraint on $E||A(X)||_2^2$, and assume that the noise variance components can be arranged in ascending order. This can only be done if the channel is finite-dimensional. In that case, one can take $R_W = I$, the identity, and thereby use a true power constraint. The assumption (A-1) becomes $R_N = I + S$, and the capacity is as given in (a); this agrees with the referenced results in [4].

[1] C. R. Baker, Capacity of the Gaussian channel without feedback, Information and Control, 37, 70-89 (1978).

[2] C. R. Baker, Channel models and their capacity, in "Essays in Statistics: Contributions in Honor of Norman L. Johnson", P.K. Sen, ed., North-Holland, New York (1983).

[3] P. M. Ebert, The capacity of the Gaussian channel with feedback, Bell System Tech. J., 49, 1705-1712 (1970) .

[4] R. M. Fano, "Transmission of Information", M.I.T. Press, Cambridge, and Wiley, New York (1961).

[5] R. G. Gallager, "Information Theory and Reliable Communication", Wiley, New York (1968).

[6] C. R. Rao and V. S. Varadarajan, Discrimination of Gaussian processes, Sankhyā, Series A, 25, 303-330 (1963).

[7] G. R. Walsh, "Methods of Optimization", Wiley, New York (1975).